第 34 卷第 3 期
2022 年 3 月

强 激 光 与 粒 子 束
HIGH POWER LASER AND PARTICLE BEAMS

Vol. 34, No. 3
Mar., 2022

·激光前沿交叉科学·

面向嵌入式平台的轻量化神经网络手势识别方法*

杨晨奕， 何玉青， 赵俊媛， 李国荣

（北京理工大学 光电学院, 光电成像技术与系统教育部重点实验室, 北京 100081）

摘 要： 针对传统基于图像分割和特征提取的手势识别算法在复杂背景下识别准确率低、灵活性差的问题，基于目标检测神经网络的手势识别算法可以有效提高复杂环境下手势识别的准确性。受嵌入式处理器体积和功耗的限制，常用的目标检测神经网络在嵌入式上的识别速度较低，不能满足实时手势识别的要求。在SSD目标检测的基础上对其进行优化，使用MobileNetv3网络实现特征提取，目标检测方面则是使用SSD-lite结构，其使用深度可分离卷积替代普通卷积，实现了轻量化MobileNetv3-SSDLite手势识别算法的设计。针对手势识别的要求，制作了包含不同手势的数据集，利用它在服务器上完成了模型的训练。为了满足嵌入式的算力限制，通过模型的量化压缩将float64的网络参数量化为int8，并压缩网络结构，提高网络在嵌入式上的推理速度，实现基于嵌入式的手势识别。实验结果表明，基于嵌入式的MobileNetv3-SSDLite手势识别算法可以达到平均准确率99.61%，且识别速度达到每秒50帧以上，满足实时手势识别的要求。

关 键 词： 手势识别；深度神经网络；嵌入式；轻量化；MobileNev3-SSDLite

中图分类号： TP391 **文献标志码：** A **doi**: 10.11884/HPLPB202234.210335

Lightweight neural network hand gesture recognition method for embedded platforms

Yang Chenyi， He Yuqing， Zhao Junyuan， Li Guorong

（*Key Laboratory of Photoelectronic Imaging Technology and System of Ministry of Education, School of Optics and Photonics, Beijing Institute of Technology, Beijing* 100081, *China*）

Abstract: Compared with the traditional gesture recognition algorithms based on image segmentation and feature extraction in complex backgrounds which have low recognition accuracy and poor flexibility, the gesture recognition algorithm based on target detection neural network can effectively improve the accuracy of gesture recognition in complex environments. Restricted by the size and power consumption of embedded processors, the recognition speed of commonly used target detection neural networks on embedded processors is low and cannot meet the requirements of real-time gesture recognition. In this paper, we optimize the SSD target detection and use MobileNetv3 network to achieve feature extraction and SSD-lite structure for target detection, thus to use depth-separable convolution instead of ordinary convolution to realize the design of lightweight MobileNetv3-SSDLite gesture recognition algorithm. For the requirements of gesture recognition, we make a dataset containing different gestures and complete the training of the model on the server using the dataset. In order to meet the arithmetic limitation of embedded processor, we quantize the float64 network parameters into int8 by quantization compression of the model, and compress the network structure to improve the inference speed of the network on embedded processor to realize the embedded-based gesture recognition. The experimental results show that the embedded-based MobileNetv3-SSDLite gesture recognition algorithm can achieve an average accuracy of 99.61% and a recognition speed of above 50 frame/s, which meets the requirements of real-time gesture recognition.

Key words: hand gesture recognition, deep neuron network, embedded system, lightweight, MobileNetv3-SSDLite

* **收稿日期:**2021-07-30； **修订日期:**2021-12-21
基金项目:国家重点研发计划项目（2020YFF0304104）
联系方式:杨晨奕, 3120190590@bit.edu.cn。
通信作者:何玉青, yuqinghe@bit.edu.cn。

手势控制是人机交互方式之一，通过判断用户的手部位置和手势姿态来实现相应的控制操作[1]。相比于传统通过鼠标、轨迹球和键盘的交互方式，通过用户手势可以实现更为自然、人性化的人机交互[2]。目前手势识别主要分为接触式和非接触式，其中非接触手势识别是通过超声波，单目、双目摄像机或深度摄像机捕捉用户的手部位置和姿态，实现手势识别，进而实现对其他设备的非接触控制。相较于需要穿戴检测手套的接触式手势识别，非接触式手势识别不需要专门的穿戴设备，可以提供更好的用户体验。但另一方面，非接触式手势识别需要先判断用户的手部位置，分割并识别用户的手势信息，这也为手势识别的速度和准确性带来了一定的难度。

目前的手势识别算法主要有基于双目摄像机、RGB-D 摄像机的深度图像手势识别[3-4]和基于单目摄像机的平面手势识别。相比于深度图像手势识别，平面图像手势识别的硬件需求更小，可以直接在现有硬件基础上进行改进。但另一方面，由于缺少了图像的深度信息，基于平面图像的手势识别需要更高效的特征提取，进而给特征提取和手势识别的程序设计增加了难度。基于单目相机的手势识别可以通过传统颜色空间图像实现手部的分割[5]，利用 HOG 特征[6]、SVM 向量机[7]等方式对分割图像中的手势进行分类，得到最终的手势识别结果。随着深度学习神经网络的发展，通过神经网络实现手势的分割与检测已经成为新的趋势，深度学习神经网络可以用于对分割后的手势进行分类识别，如基于改进 AlexNet 与随机森林的手势分类神经网络和基于 VGG16 图像分类神经网络的手势分类器[8-9]。除此之外，深度学习神经网络也可以直接用于用户手部的分割和识别，通过端到端的方式直接实现最终的手势识别功能，目前已有基于 YOLO 与 SSD 目标检测神经网络的手势识别算法，应用于手语翻译等领域[10-12]。相比于传统的图像分割-特征提取-分类的流程，基于深度学习神经网络的手势识别不需要人工设置特征，卷积神经网络可以自行提取特征，完成手势的分割与识别。因此基于深度学习神经网络的手势识别算法可以很方便地修改网络结构，灵活设置不同的控制手势，提高网络的搭建效率和灵活性。不过现在大部分基于深度学习神经网络的手势识别算法的算力要求较高，因而大部分的算法需要在服务器或者 PC 上运行，不适合算法的应用和部署。而嵌入式处理器作为一种小体积、低功耗的边缘计算设备，可以较好地在单目摄像机硬件上实现改装，实现低功耗、高效率的手势识别控制[13]。

因此，本文提出了一种基于嵌入式单目相机的手势识别系统。搭载在嵌入式处理器上的 MobilNet-SSDLite 手势识别神经网络接收摄像机输出的视频帧，通过神经网络的推理得到识别的用户手势，根据用户的不同手势控制信息以实现进一步的操作。相比于需要专门硬件的双目摄像机手势识别和 RGB-D 相机的手势识别，本文所提出的基于嵌入式的静态手势识别系统可以在现有摄像机的基础上实现，无需专门的硬件系统，并且嵌入式处理器体积小、功耗低，可以长时间可靠工作，提高了手势识别系统的稳定性。该系统可以实现设备的远程非接触控制，如控制录像设备的启停及视频截图等，也可用于安全设备的手部侵入告警，提高工作的便利性和安全性。

1 基于改进 SSD 网络实现轻量化手势识别算法

手势识别深度学习网络是整个手势识别算法的核心，该网络接收摄像机捕捉的视频帧，利用神经网络算法处理视频帧并输出手势的类别和位置。相比于其他手势识别算法，基于深度学习神经网络的手势识别算法可以自主完成手势特征的提取和分析，提高了手势识别算法的灵活性，降低了设计难度。

本文所使用的手势检测深度学习网络是在 SSD 深度学习目标检测网络[14]的基础上进行轻量化改进得到。相比于 Faster-RCNN[15]，YOLO[16]等其他目标检测网络，SSD 网络在 COCO 目标检测数据集上的目标检测精度较高，并且 SSD 网络的目标检测推理速度也较快。但是，考虑到嵌入式处理器的计算能力，SSD 目标检测网络仍不能满足实时手势检测任务，所以，为了搭建面向嵌入式平台的轻量化手势识别算法，需要在 SSD 网络的基础上进行改进，以实现轻量化 MobileNetv3-SSDLite 手势识别算法的搭建。算法的搭建及部署流程框图如图 1 所示。

在图 1 中，网络的训练和优化是在电脑端完成，经过训练和压缩优化之后的手势识别网络部署到嵌入式处理器中，部署完成的手势识别网络在嵌入式处理器上推理并给出手势识别的结果。

1.1 特征提取网络的轻量化

文献 [14] 中 SSD 使用的特征提取网络为 VGG16 网络[17]，该网络是 2014 年 ImageNet 图像分类大赛的第二名，有较好的图像特征提取能力。但是 VGG16 网络的参数量较大，对处理器的运算能力要求较高，难以在嵌入式处理器上实现实时手势识别。而简单的减少 VGG16 的网络层数，又会影响网络的特征提取能力，降低最终的手势识别精度。因此，在不影响特征提取网络效果的前提下降低网络的参数和运算量，需要一种有效的轻量化卷积操作和高效的网络设计。

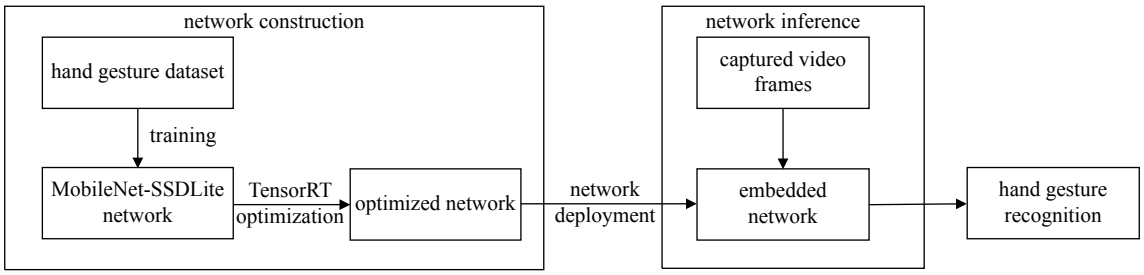


Fig. 1　Construction and pipeline of the algorithm

图 1　算法搭建及工作流程框图

1.1.1　深度可分离卷积

由于考虑到 VGG16 网络的上述问题，本文采用基于深度可分离卷积思路设计的 MobileNetv3 网络[18] 作为手势识别的特征提取网络。深度可分离卷积（Depthwise Convolution）是一种高效的轻量化卷积思路[19]，不同于传统卷积，它将传统卷积中“卷积+通道调节”的计算分为两次进行，减少了卷积过程中的卷积核数量和运算量，实现了卷积运算的轻量化。

传统卷积神经网络中，对于一个输入通道数为M，输出通道数为N的卷积层，如果输入的上一层特征图尺寸为$H\times W$，卷积核尺寸为$K\times K$。则卷积过程需要N组卷积核，其中每组卷积核的通道数为M，输入特征图的每个通道分别与每一组卷积核中的卷积核进行卷积，再将每组通道得到的卷积结果求和，也就得到了输出特征图中的一个通道。对N组卷积核重复以上的操作，就得到了最终的卷积输出结果。整个卷积操作的运算量为HWK^2MN。通常来说，特征图的通道数量都在 10^2～10^3 数量级，所以传统卷积操作需要较大的运算量。

而深度可分离卷积则是将卷积分为两步进行。如图 2 所示，同样对于一个输入通道数为M，输出通道数为N，尺寸为$H\times W$特征图。深度可分离卷积的第一步是使用与输入特征图通道数相同的卷积核进行卷积，获得卷积后的特征图，其中卷积后的特征图通道数与输入特征图的通道数保持不变；第二步是利用N组通道数为M的1×1卷积调整卷积后特征图的通道数，使最终得到的输出特征图通道数为N。卷积与通道调整的运算量分别为HWK^2M与$HWMN$，深度可分离卷积与传统卷积的运算量之比为

$$\frac{HWK^2M+HWMN}{HWK^2MN}=\frac{1}{N}+\frac{1}{K^2} \tag{1}$$

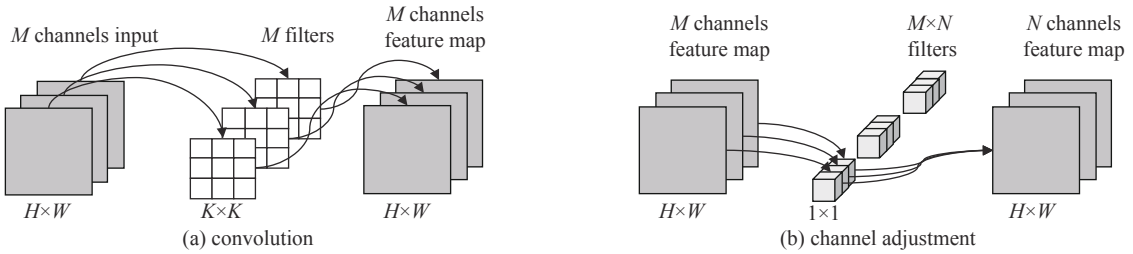


Fig. 2　Depthwise separable convolution

图 2　深度可分离卷积

通常来说，对于3×3尺寸的卷积核，针对 10^2～10^3 数量级的特征图，使用深度可分离卷积的运算量理论上可以达到使用传统卷积的1/8 ~ 1/9。

1.1.2　通道注意力机制

受限于卷积核的数量，深度可分离卷积的特征提取效果要弱于传统卷积的，所以本文所使用的 MobileNetv3 特征提取网络使用深度可分离卷积，继承上一代网络的倒残差结构[20] 以外，还使用了轻量化的特征图压缩注意力机制[21]，在图像卷积过程中对不同的特征图赋予不同的权重，使得更重要的特征得以保留，提高网络的特征提取效果。压缩注意力机制结构如图 3 所示。

在图 3 中，特征图尺寸为$H\times W$，通道数为C，r为编码压缩后的通道注意力参数。经过全局池化（global pooling）、全连接网络和 sigmoid 激活函数后，特征图中每一层有了不同的权重，再与之前的特征图逐层相乘，便获得了带权重参数的特征图。相比于无通道注意力机制的 MobileNet，使用了通道注意力机制之后，运算量增加了 0.5%，ImageNet

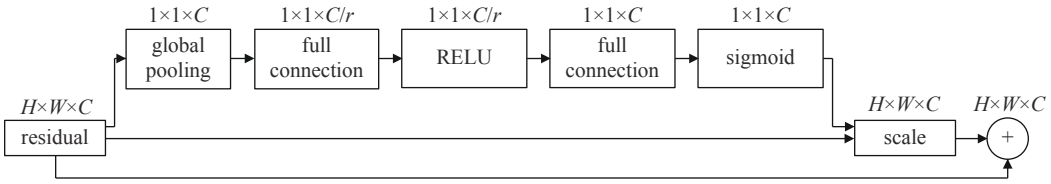


Fig. 3 Squeeze-and-excitation module[22]

图 3 压缩注意力机制[22]

图像分类任务的错误率降低了 3%, 可见通道注意力机制在基本不改变特征提取网络运算量的同时, 有效提高了网络的特征提取效果。

1.1.3 应用 MobileNetv3 实现手势特征提取网络轻量化

本文使用 MobileNet 特征提取网络系列的第三代 MobileNetv3 作为手势特征提取网络, 从输入的图像中提取特征, 生成手势特征图并输入到深层目标检测网络中。MobileNet系列网络的参数量、乘加计算量(MAC)和 ImageNet 图像分类数据集准确率与 VGG16 的比较如表 1 所示。

从表 1 中可以看出, Mobilev3 骨干网络的参数量和运算量都远小于传统 VGG16 骨干网络的, 且其在 ImageNet 图像分类数据集中的分类准确率也要高于 VGG16 网络的。因此使用 MobileNetv3 网络用于手势识别的特征提取网络可以在不影响手势特征提取能力的前提下有效降低网络的参数量和运算量, 提高最终手势识别算法的速度。

不过文献 [18] 中 MobileNetv3 被用于实现图像分类, 网络的最后几层为完成图像分类的全连接神经网络, 并非生成用于目标检测的特征图; 另一方面, SSD 目标检测网络的深层检测部分采用的是多尺度特征图目标检测的策略, 所以要实现基于 MobileNetv3 的轻量化手势特征提取网络, 需要在原有 MobileNetv3 的基础上修改最后输出的特征图尺寸和通道数, 使得手势识别神经网络的特征提取网络与深层目标检测网络相匹配。

表 1 MobileNet 系列与 VGG16 的对比

Table 1 MobileNet series comparison to VGG16

network structure	params/Mbyte	MACs/10^6	ImageNet accuracy/%
VGG16	13.8	15 300	71.5
MobieNetv1	4.2	569	70.6
MobileNetv2	3.4	300	72.0
MobileNetv3	5.4	219	75.2

因此, 本文使用的 MobileNet 在最后一层卷积层以外, 还添加了额外的卷积层用于调整输出的特征图尺寸和通道数。一方面使得特征提取网络的输出与深层 SSD 目标检测网络的输入相匹配, 另一方面也加深了网络的深度, 在一定程度上提高了特征提取网络的能力。生成额外卷积层的操作也使用了深度可分离卷积, 压缩注意力机制等轻量化操作。输出用于检测的特征图及其尺寸如表 2 所示。

表 2 用于检测的额外特征图及其尺寸

Table 2 Extra feature map layers for object detection

extra layers	shape
layer 1	39×39×512
layer 2	19×19×1024
layer 3	10×10×512
layer 4	5×5×256
layer 5	3×3×256
layer 6	1×1×256

1.2 深层目标检测网络的轻量化

深层目标检测网络接收特征提取网络输出的特征图, 通过分类与回归的方式完成目标检测任务。SSD 深层目标检测网络同时在大小不同的特征图上设置一系列的先验框, 判断先验框与数据集中给出手势真实检测框的交并比 IoU, 进而分析特征图中给出的特征中是否存在手势及手势类别。如表 3 所示, SSD 目标检测深层网络同时在 6 个大小不同的特征图上检测, 其中尺寸较小的特征图用于检测较大(较近)的手势, 尺寸较大的特征图用于检测较小(较远)的目标。考虑到重复检测的问题, SSD 网络使用非极大值抑制的方式删除检测效果较差的检测框, 仅保留置信度最高的结果。

表 3 SSDLite 深层检测网络与 SSD 的对比

Table 3 SSDLite detection head comparison to SSD

network structure	params/Mbyte	MACs/10^6	mAP/%
SSD	14.8	1250	19.3
SSDLite	2.1	350	22.2

从表 3 中可以看出, 使用轻量化的 SSDLite 轻量化深层目标检测网络, 可以有效降低网络的参数量和运算量, 并且将所有卷积层替换为深度可分离卷积的轻量化操作对目标检测的结果影响也十分有限。所以 SSDLite 轻量化深层目标检测网络适合于嵌入式实时手势识别任务。本文提出的轻量化手势识别算法结构如图 4 所示。

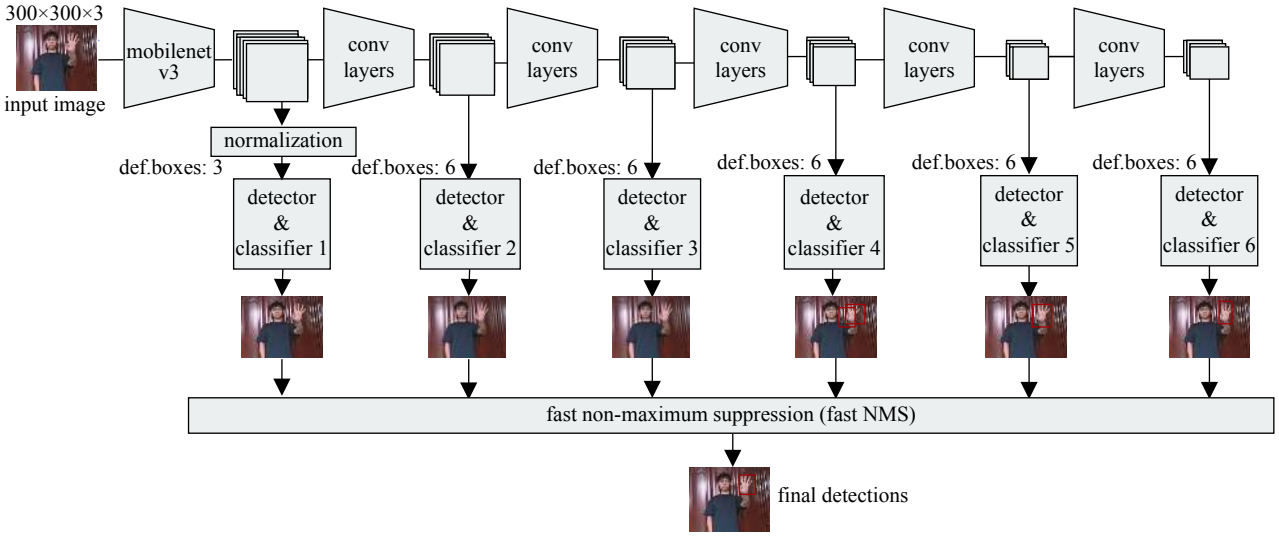


Fig. 4 Hand gesture recognition network based on MobileNetv3-SSDLite

图 4 基于 MobileNetv3-SSDLite 的手势识别算法

从图 4 中可以看出，输入图像为长宽各为 300 像素的彩色图像，经过 MobileNetv3 提取特征并由额外卷积层生成表 2 所示的特征图用于 SSDLite 深层目标检测网络的手势识别。不同尺寸特征图的手势识别结果由非极大值抑制模块筛选，得到最终的手势识别结果。

1.3 手势识别算法在嵌入式上的部署

搭建并获得了手势识别算法模型之后，需要在服务器上使用手势识别数据集对算法进行训练，以获得最终训练完成的手势识别算法。为了得到最好的训练效果，训练过程中所使用的模型参数都是全精度参数。一方面，为了保证算法模型搭建的精确度，模型每一层都独立存在，这样的优点是可以获得最完整灵活的算法结构以及最高精度的训练结果。另一方面，训练完成的手势识别算法如果直接在嵌入式处理器上运行，则会受制于全精度模型参数的运算和模型结构的加载，导致最终在嵌入式上手势识别算法推理速度的降低。因此，为了提高手势识别算法在嵌入式处理器上的推理速度，在算法迁移至嵌入式处理器之前需要先对模型量化和压缩，得到最优的嵌入式推理速度。

模型量化是指将算法模型中使用 32 位浮点数保存的模型参数量化至 16 位整形数甚至 8 位整形数，进而降低嵌入式处理器的运算压力和数据吞吐压力。量化前后网络参数的对应关系为[22]

$$S=\frac{r_{\max}-r_{\min}}{q_{\max}-q_{\min}} \tag{2}$$

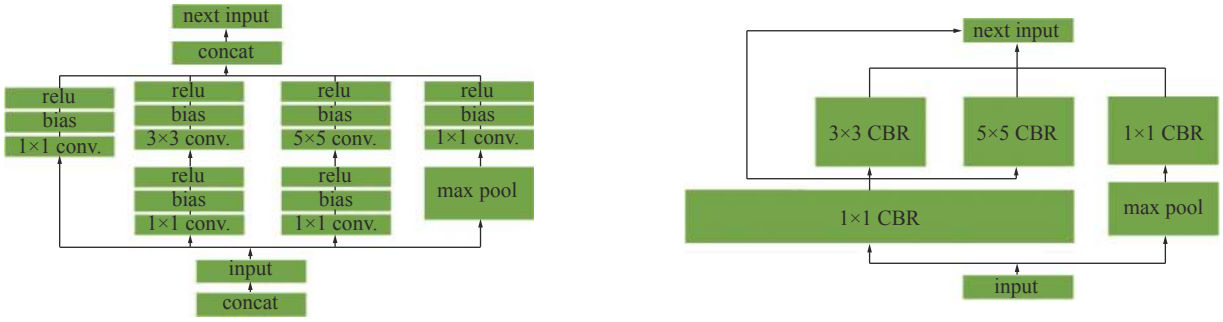
$$Z=\mathrm{round}(q_{\max}-q_{\min}) \tag{3}$$

$$q=\mathrm{round}\left(\frac{r}{S}+Z\right) \tag{4}$$

式中：r为训练得到的浮点参数；q为量化后得到的整形参数；S和Z分别为参数量化过程中的缩放因子和偏移量。

假定训练得到的网络参数为 32 位浮点数，经过量化得到 16 位整数的网络参数，如式(2)所示，首先根据网络训练得到的参数$r_{\max}$，$r_{\min}$，然后量化需要的$q_{\min}$和$q_{\max}$计算得到缩放因子S，再由式(3)计算出网络参数的偏移量Z，最后，需要的量化参数q根据之前获得的缩放因子和偏移量即可得到。不过式(3)和式(4)的计算过程中存在两次取整操作，在取整过程中会出现一定的精度损失。所以在量化过程中需要仔细设置最终量化参数的大小，在模型的推理速度与推理精度之间取得平均。

除了算法模型的参数量化，另一个模型的优化方案是模型融合，即将卷积层(Conv Layer)、偏置层(Bias Layer)和激活函数层(Activate Layer)融合成一层，一次性读取所有参数，降低嵌入式处理器的数据 IO 瓶颈。除此之外，也可以将网络水平组合，合并相同张量和操作相同的网络层，进一步提高网络在嵌入式上的推理速度。优化前后的神经网络结构如图 5 所示，优化前的神经网络如图 5(a)所示，可以看出此时的神经网络结构较为复杂，嵌入式神经网络推理需要较多的数据 IO 操作；经过优化后的神经网络结构如图 5(b)所示，此时卷积-偏置-激活被压缩到一



(a) original network　　(b) optimized network

Fig. 5　Neural network structure before and after the embedded optimization

图 5　嵌入式优化前与优化后的神经网络结构图

个 CBR 层中，且操作相同的网络得到了合并，一些冗余层也被移除。因此，比起原始的神经网络，经过嵌入式推理优化的神经网络结构更为简洁，适合于嵌入式的神经网络推理。

2　实　验

2.1　基于 MobileNetv3-SSDLite 手势识别算法的训练、测试和部署

不同于常见的目标检测，手势识别没有一个标准的手势数据集，目前大部分手势识别数据集都是针对自己的手势识别任务而建立的，如手语识别、指尖识别等，且绝大部分手势识别数据集都只有手部图片，无法用于手势识别算法的训练。因此我们模拟实际的应用场景，制作静态手势识别数据集。本文选取了如图 6 所示 5 种指令手势，分别为握拳、食指伸出、ok、大拇指收回、五指伸出，根据伸出的手指数编号为手势 0、手势 1、手势 3、手势 4、手势 5。之所以没有选用深处食指和中指，摆出“V”字形的手势，是因为该手势出现的频率较高，如果将其设置为识别手势的话可能会与用户的想法发生冲突，因此不选取其作为识别手势。

Fig. 6　The chosen hand gestures

图 6　选取的手势示意图

用手机相机拍摄包含手势指令的动作视频，每 8 帧抽取视频帧图像作为原始数据。从中挑选手势清晰、包含左右手、多角度、不同距离的手势图像作为可用数据集，如图 7(a)所示。对于一些出现运动模糊、手势角度过大无法辨别、图像失真或手势部分超出画面等情况的帧图像，如图 7(b)所示，对其舍弃。对选取的可用手势图像或背景图，使用 labelImg 工具进行手势标注，得到可用的 xml 文件，它指明了图像的文件名、路径、图像大小、手势分类以及手势位置等信息，供模型训练测试使用。最终得到每个手势图像 2100 张，包含背景图在内的数据集图像共 12 000 张，图像尺寸大小共三类：1280×720, 960×544, 1920×1080 像素，且数量均匀。

(a) images suitable for dataset　　(b) images unsuitable for dataset

Fig. 7　Selecting images for hand gesture dataset

图 7　手势训练集示意图

获得数据集之后，将其根据 80%～20% 比例划分为训练集和验证集，其中训练集图像 9600 张，验证集图像 2400 张。利用训练集在服务器(Intel Xeon 4110 CPU, 256G RAM, NVIDIA GV100 GPU)上训练 MobileNetv3-SSDLite

算法模型，训练的回归损失、分类损失和平均损失随训练步数 epoch 的变化曲线如图 8 所示。

从图 8 中可以看出，损失函数值随着神经网络训练的进行不断降低，并在训练步数 120 后收敛。

为了验证手势识别算法的识别准确率，使用验证集 2400 张手势图像测试训练得到的手势识别算法。验证集的手势图像包含室内简单背景、光照充足、单人单个手势的情况，识别结果如表 4 所示。

从表 4 中可以看出，所有手势的识别准确率都在 99% 以上，其中手势 4 的识别准确率最低，为 99.22%；手势 1 的识别准确率最高，为 100.00%，平均准确率为 99.61%。其中，不同手势识别准确率不同是因为手势的复杂度不同，深度学习神经网络在提取特征与目标检测过程中会出现特征提取或者分析错误的情况。如手势 4 与手势 5 的相似度较高，在手势的特征提取与识别过程中受到背景干扰，出现手势识别的混淆。

为了进一步测试手势识别算法在不同场景下的手势识别准确率，测试了多个手势，复杂背景以及光照不足的情况，平均手势识别准确率如表 5 所示。

从表 5 中可以看出，本文所提出的 MobileNetv3-SSDLite 手势识别算法在场景中存在多个手势时识别准确率出现了一定的降低，与表 4 中的识别结果区别不大，这是因为多个手势的场景与训练用数据集的场景比较接近，手势识别算法可以较好地应用数据集中学习得到的手势特征。手势识别精度最低的场景是复杂背景下的手势识别，这个场景中存在数个人，其中有些人做手势而另外的人处于自然状态。这个场景的识别精度低是出现了较多的误检情况，把处于自然状态下人的手检测为手势。相反，光照不足的场景下则是出现了较多的漏检，一些手势未能准确识别。综合分析上面 3 种不同的场景，识别率较低的均为数据集中未出现的场景，而与数据集场景重合度较高的多个手势场景则手势识别准确率相对较好。因此，对于数据集涵盖的场景，本文提出的 MobileNetv3-SSDLite 手势识别算法的识别效果较好，而数据集未涵盖的场景，则识别精度较低。所以在未来的应用过程中，应当根据实际的使用场景选择并制作合适的数据集。

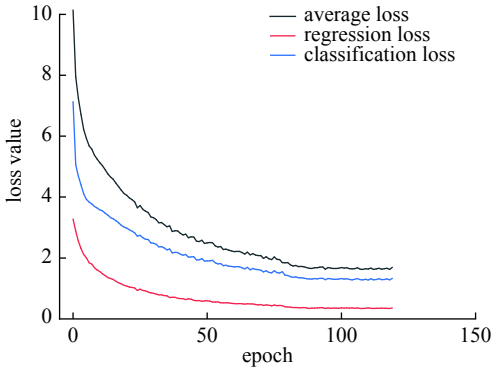


Fig. 8　Network loss in training process

图 8　训练中的损失函数变化情况

表 4　不同手势的识别结果

Table 4　Recognition results of hand gestures

hand gesture	accuracy/%
0	99.64
1	100.00
3	99.51
4	99.22
5	99.69
average	99.61

表 5　不同场景下手势识别结果

Table 5　Recognition results of hand gestureson various scenarios

scenarios	average accuracy/%
multiple hand gestures	96
complicated background	64
low light intensity	72

Fig. 9　NVIDIA Jetson TX2 embedded processor developer kit

图 9　NVIDIA Jetson TX2 嵌入式处理器开发者套件

如图 9 所示，将训练好的算法模型通过 TensorRT 实现参数量化和模型压缩，并部署到 NVIDIA Jetson TX2 处理器上，该嵌入式处理器有专门用于神经网络的运算核心，以 15 W 的功率提供 1.33TFLOPS 的算力，也可以选择 7.5 W 的低功耗模式，适合于本文中静态手势识别算法的部署。

使用图 9 中的 TX2 开发者套件进行测试，TX2 嵌入式处理器读取开发者套件上 CSI 摄像机的视频帧，使用搭载在 TX2 上的 MobileNetv3-SSDLite 网络推理实现手势识别，部分检测识别结果如图 10 所示。

由图 10 可以看出，搭载在 TX2 嵌入式处理器上的 MobileNetv3-SSDLite 手势识别算法可以有效定位用户手势，给出准确的手势识别结果。

2.2　与其他算法的比较

为了验证本文提出的基于 MobileNetv3-SSDLite 手势识别算法的轻量化设计，使用相同的手势数据集训练了

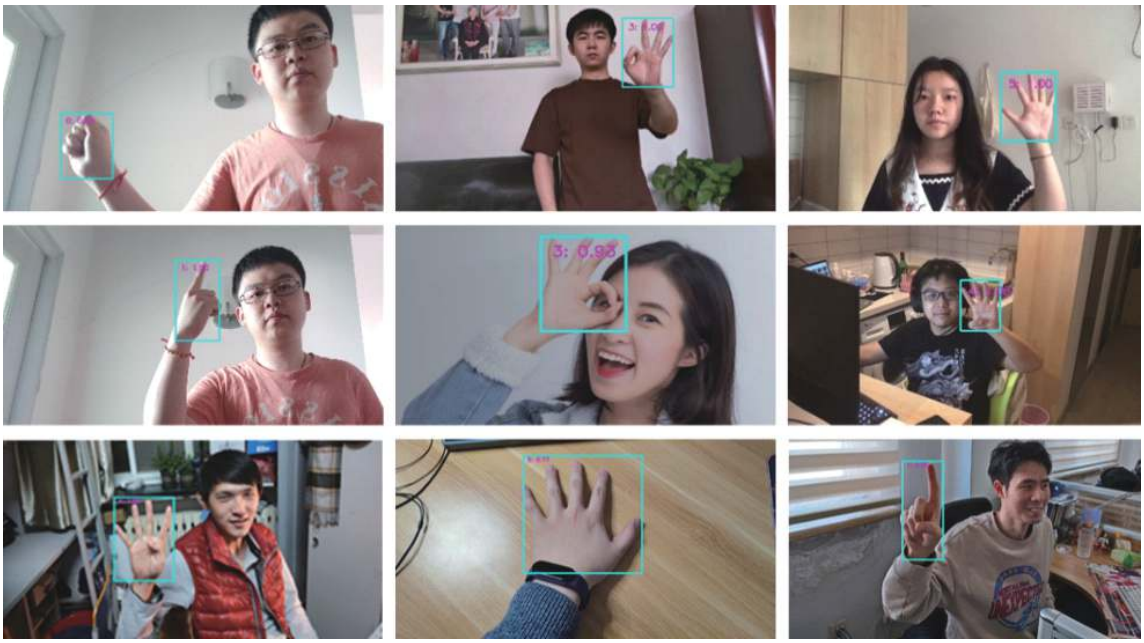


Fig. 10 Part of the hand gesture recognition results

图 10 部分手势识别结果

4 个 VGG16-SSD, MobileNetv1-SSD, MobileNetv1-SSDLite 和 MobileNetv2-SSDLite 手势识别算法。不同算法的参数量、速度和验证集上手势的平均识别准确率如表 6 所示。

表 6 不同手势识别算法的比较

Table 6 Comparison of different hand gesture recognition algorithms.

algorithm	params/Mbyte	MACs/10^6	frame rate/(frame/s)	mean accuracy/%
VGG16-SSD	24.3	30654	2	91.75
MobieNetv1-SSD	7.2	1299	12	93.98
MobileNetv1-SSDLite	4.1	1130	16	93.86
MobileNetv2-SSDLite	3.1	656	36	91.01
MobileNetv3-SSDLite	2.2	526	58	99.61

由表 6 可以看出，相比于最初提出的 VGG16-SSD 模型和其他 MobileNet-SSD 的模型，本文提出的 MobileNetv3-SSDLite 手势识别模型的参数量最少，需要的算力最低，识别速度可以达到 58 帧/s，满足实时性要求，最终的手势识别准确率最高。

3 结 论

本文提出了一种基于嵌入式 MobileNetv3-SSDLite 的手势识别算法。根据设计的识别手势，制作了手势识别数据集，搭建基于嵌入式的 MobileNetv3-SSDLite 算法，在服务器上训练算法模型并利用 TensorRT 量化和压缩模型，部署到 NVIDIA Jetson TX2 嵌入式处理器上。

在嵌入式处理器上测试本文提出的基于 MobileNetv3-SSDLite 手势识别算法。经过测试，本文所提出的基于嵌入式 MobileNetv3-SSDLite 的手势识别算法准确率较高，在嵌入式处理器上实现 50 帧/s 以上的手势识别帧率，满足实时手势识别的要求。

参考文献：

[1] 陈壮炼, 林晓乐, 王家伟, 等. 基于卷积神经网络的手势识别人机交互系统的设计[J]. 现代计算机, 2021(6): 57-62.（Chen Zhuanglian, Lin Xiaole, Wang Jiawei, et al. Design of human-computer interaction system for gesture recognition based on convolutional neural network[J]. Modern Computer, 2021(6): 57-62）

[2] 袁博, 查晨东. 手势识别技术发展现状与展望[J]. 科学技术创新, 2018(32): 95-96.（Yuan Bo, Zha Chendong. Gesture recognition technology development status and outlook[J]. Scientific and Technological Innovation, 2018(32): 95-96）

[3] 时梦丽, 张备伟, 刘光徽. 基于深度图像的实时手势识别方法[J]. 计算机工程与设计, 2020, 41(7): 2057-2062.（Shi Mengli, Zhang Beiwei, Liu Guanghui. Real-time gesture recognition method based on depth image[J]. Computer Engineering and Design, 2020, 41(7): 2057-2062）

[4] 彭理仁, 王进, 林旭军, 等. 一种基于深度图像的静态手势神经网络识别方法[J]. 自动化与仪器仪表, 2020(1): 6-9,15.（Peng Liren, Wang Jin, Lin Xujun, et al. A static gesture recognition method based on depth image and neural network[J]. Automation & Instrumentation, 2020(1): 6-9,15）

[5] 吴铁凡, 郭剑辉. 一种基于肤色模型的改进型手势分割算法的实现[J]. 电子设计工程, 2020, 28(18): 185-188,193.（Wu Yifan, Guo Jianhui. Implementation of an improved gesture segmentation algorithm based on skin color model[J]. Electronic Design Engineering, 2020, 28(18): 185-188,193）

[6] Li Hui, Yang Lei, Wu Xiaoyu, et al. Static hand gesture recognition based on HOG with Kinect[C]//Proceedings of the 2012 4th International Conference on Intelligent Human-Machine Systems and Cybernetics. 2012: 271-273.

[7] Liua C, Zhou Shuwang, Hu Sheng, et al. Hand gesture recognition based on sEMG signal and improved SVM voting method[C]//Proceedings of the 2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE). 2020: 605-608.

[8] 石雨鑫, 邓洪敏, 郭伟林. 基于混合卷积神经网络的静态手势识别[J]. 计算机科学, 2019, 46(s1): 165-168.（Shi Yuxin, Deng Hongmin, Guo Weilin. Static gesture recognition based on hybrid convolution neural network[J]. Computer Science, 2019, 46(s1): 165-168）

[9] Hussain S, Saxena R, Han Xie, et al. Hand gesture recognition using deep learning[C]//Proceedings of the 2017 International SoC Design Conference (ISOCC). 2017: 48-49.

[10] 郭紫嫣, 韩慧妍, 何黎刚, 等. 基于改进的YOLOV4的手势识别算法及其应用[J]. 中北大学学报(自然科学版), 2021, 42(3): 223-231.（Guo Ziyan, Han Huiyan, He Ligang, et al. Gesture recognition algorithm and application based on improved YOLOV4[J]. Journal of North University of China (Natural Science Edition), 2021, 42(3): 223-231）

[11] Chhajed R R, Parmar K P, Pandya M D, et al. Messaging and video calling application for specially abled people using hand gesture recognition[C]//Proceedings of the 2021 6th International Conference for Convergence in Technology (I2CT). 2021: 1-4.

[12] Yi Chengming, Zhou Liguang, Wang Zhixiang, et al. Long-range hand gesture recognition with joint SSD network[C]//Proceedings of the 2018 IEEE International Conference on Robotics and Biomimetics (ROBIO). 2018: 1959-1963.

[13] 孔维刚, 李文婧, 王秋艳, 等. 基于改进YOLOv4算法的轻量化网络设计与实现[J/OL]. 计算机工程, 1-10(2021-04-30).（Kong Weigang, Li Wenjing, Wang Qiuyan, et al. Design and implementation of lightweight network based on YOLOv4 algorithm[J/OL]. Computer Engineering, 1-10(2021-04-30). https://doi.org/10.19678/j.issn.1000-3428.0060948）

[14] Liu Wei, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[C]//Proceedings of the 14th European Conference on Computer Vision. 2016: 21-37.

[15] Ren Shaoqing, He Kaiming, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[16] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 779-788.

[17] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[DB/OL]. arXiv preprint arXiv: 1409.1556, 2014.

[18] Howard A, Sandler M, Chen Bo, et al. Searching for MobileNetV3[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). 2019: 1314-1324.

[19] Howard A G, Zhu Menglong, Chen Bo, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[DB/OL]. arXiv preprint arXiv: 1704.04861, 2017.

[20] Sandler M, Howard A, Zhu Menglong, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018: 4510-4520.

[21] Hu Jie, Shen Li, Albanie S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.

[22] 杨国威, 许志旺, 房臣, 等. 融合剪枝与量化的目标检测网络压缩方法[J/OL]. 计算机工程与应用, 1-12[2021-12-17].（Yang Guowei, Xu Zhiwang, Fang Chen, et al. Object detection network compression method based on pruning and quantization[J/OL]. Computer Engineering and Applications, 1-12[2021-12-17]. http://kns.cnki.net/kcms/detail/11.2127.tp.20210918.1121.008.html）